A Novel Single-Molecule Study To Determine Protein−Protein Association Constants

Glenn C. Ratcliff and Dorothy A. Erie*

Contribution from the Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received November 2, 2000. Revised Manuscript Received April 30, 2001

Abstract: Atomic force microscopy (AFM) is traditionally used as an imaging technique to gain qualitative information for a biological system. We have successfully used the imaging capabilities of the AFM to determine protein−protein association constants. We have developed a method to measure the molecular weight of a protein based on its volume determined from AFM images. Our volume determination method allows for rapid, accurate analysis of large protein populations. On the basis of the measured volume, the fraction of monomers as dimers was determined for the DNA helicase UvrD, and the dissociation constant (K_d) for the helicase was calculated. We determined a K_d for UvrD of 1.4 μM, which is in good agreement with published K_d data obtained from analytical ultracentrifugation (AUC) studies. Our method provides a rapid method for determining protein−protein association constants.

Introduction

Essentially all regulatory processes within a cell involve protein−protein interactions. Various methods have been developed to study the thermodynamics of these interactions.[1] These methods include analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), and a variety of spectrophotometric techniques. Despite being reliable and accurate, these methods have limitations. For example, high concentrations of sample are required for AUC and ITC, limiting their capability for measuring tight binding interactions. Spectrophotometric techniques require a change in absorption or emission upon association. The use of SPR can be limited by nonspecific binding of proteins to its detection surface. In this paper, we present a volume analysis technique utilizing atomic force microscopy (AFM) to study protein−protein interactions.

In 1986, Binnig, Quate, and Gerber invented the atomic force microscope (AFM) to obtain atomic resolution of nonconductive surfaces.[2] The AFM has evolved into a powerful tool for studying biological systems because of its facility for imaging soft samples both in air and under solution at nanometer resolution.[3−9] In addition to being a high-resolution microscope, the unique design of the AFM permits measurement of normal and lateral forces with nanoNewton resolution.[10−19] Researchers have taken advantage of this latter property to make quantitative measurements of intra- and intermolecular forces in biological systems.[10−19] On the other hand, the imaging capabilities of the AFM have not been utilized for determining thermodynamic parameters. In this paper, we demonstrate that AFM provides a rapid method for determining protein−protein association constants.

Because AFM produces topographical images, it has been possible to relate the molecular weight of a protein to its volume determined from AFM images.[20−24] Qualitative volume analysis studies have been used to show oligomerization states of proteins.[21,24] In principle, this technique can also give quantitative information such as protein−protein association constants. In this paper, we present a rapid method using volume analysis techniques to determine protein−protein association constants. We demonstrate that the volume of a protein, as measured by

(1) Phizicky, E. M.; Fields, S. *Microbiol. Rev.* **1995**, *59*, 94−123.
(2) Binnig, G.; Quate, C. F.; Gerber, C. *Phys. Rev. Lett.* **86 A.D.**, *56*, 930−933.
(3) Bustamante, C.; Rivetti, C. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 395−429.
(4) Bustamante, C.; Erie, D. A.; Keller, D. *Curr. Opin. Struct. Biol.* **1994**, *4*, 750−760.
(5) Hansma, H. G.; Laney, D. E.; Bezanilla, M.; Sinsheimer, R. L.; Hansma, P. K. *Biophys. J.* **1995**, *68*, 1672−1677.
(6) Hansma, H. G.; Hoh, J. H. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 115−139.
(7) Czajkowsky, D. M.; Shao, Z. *FEBS Lett.* **1998**, *430*, 51−54.
(8) Engel, A.; Lyubchenko, Y.; Muller, D. *Trends Cell Biol.* **1999**, *9*, 77−80.
(9) Stoffler, D.; Steinmetz, M. O.; Aebi, U. *FASEB J.* **1999**, *13* (*Suppl. 2*), S195−S200
(10) Engel, A.; Muller, D. J. *Nat. Struct. Biol.* **2000**, *7*, 715−718.
(11) Heinz, W. F.; Hoh, J. H. *Trends Biotechnol.* **1999**, *17*, 143−150.
(12) Lee, G. U.; Chrisey, L. A.; Colton, R. J. *Science* **1994**, *266*, 771−773.
(13) Fisher, T. E.; Marszalek, P. E.; Fernandez, J. M. *Nat. Struct. Biol.* **2000**, *7*, 719−724.
(14) Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature* **1998**, *393*, 181−185.
(15) Benoit, M.; Gabriel, D.; Gerisch, G.; Gaub, H. E. *Nat. Cell Biol.* **2000**, *2*, 313−317.
(16) Engel, A.; Gaub, H. E.; Muller, D. J. *Curr. Biol.* **1999**, *9*, R133−R136
(17) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−1112.
(18) Moy, V. T.; Florin, E. L.; Gaub, H. E. *Science* **1994**, *266*, 257−259.
(19) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694−3699.
(20) Henderson, R. M.; Schneider, S.; Li, Q.; Hornby, D.; White, S. J.; Oberleithner, H. *Proc Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8756−8760.
(21) Oberleithner, H.; Schneider, S.; Larmer, J.; Henderson, R. M. *Kidney Blood Pressure Res.* **1996**, *19*, 142−147.
(22) Scheuring, S.; Ringler, P.; Borgnia, M.; Stahlberg, H.; Muller, D. J.; Agre, P.; Engel, A. *EMBO J.* **1999**, *18*, 4981−4987.
(23) Schneider, S. W.; Larmer, J.; Henderson, R. M.; Oberleithner, H. *Pflugers Arch.−Eur. J. Physiol.* **1998**, *435*, 362−367.
(24) Wyman, C.; Rombel, I.; North, A. K.; Bustamante, C.; Kustu, S. *Science* **1997**, *275*, 1658−1661.

10.1021/ja005750n CCC: $20.00
Published on Web 05/26/2001

AFM, depends linearly on the molecular weight for seven different proteins ranging in molecular weight from 41 000 to 670 000. We determined the dimerization constant of the DNA helicase UvrD, which plays an essential role in both methyl-directed mismatch repair and base excision repair.[25,26]

Materials and Methods

Materials. *E. coli* UvrD helicase was a gift from Steve Matson in the Department of Biology at the University of North Carolina at Chapel Hill. The following proteins were used as molecular weight standards: Alcohol dehydrogenase, Bovine serum albumin, β-amylose, *Thermus thermophilus* RNA polymerase, Apoferritin, and thyroglobulin.

Atomic Force Microscopy Imaging. Imaging was performed with a Nanoscope IIIa instrument (Digital Instrument, Santa Barbara, CA) using Tapping mode in air. Nanosensor Pointprobe noncontact/tapping mode sensors (Molecular Imaging, Inc., Phoenix, AR) with spring constants of 48 N/m and resonance frequencies of 190 kHz were used for all imaging. The proteins were equilibrated at 37 °C for 15 min in 20 mM Tris-HCl (pH 8.3), 0.2 M NaCl, 20% glycerol, 1 mM EDTA, and 15 mM β-mercaptoethanol. The equilibrated proteins were deposited onto freshly cleaved mica (Spruce Pine Mica Co. Spruce Pine, NC), washed with deionized distilled water, and dried with a stream of N_2(g). To obtain the proper surface coverage, the deposition time varied between 5 and 60 s depending on protein concentration. All images were collected at a scan rate of 3.0 Hz and a scan size of 1 μm. This scan size was determined to give accurate volume information.

Volume Analysis. Three steps are required for determining protein volume from AFM data: (i) image planefitting, (ii) image analysis, and (iii) volume calculation. To ensure a flat background before image analysis, the image is flattened and planefitted by using the microscope's analysis software (Digital Instrument, Version 4.42r4). Image analysis is performed with the freeware program ImageSXM (based on NIH Image developed at the National Institutes of Health), which allows raw image files to be opened without losing image information.[27] (ImageSXM is available for download at http://reg.ssci.liv.ac.uk/.) The key steps for volume analysis are discussed below while a detailed description illustrating each step using ImageSXM is provided in the Supporting Information.

The first step in image analysis is to determine the height of the surface (S), which is generally nonzero. This surface height must be subtracted from the measured height of each protein before volume determination. After the surface height is measured, each protein is then highlighted individually using the density slice utility in ImageSXM. The density slice selects the pixels above the surface that represent the proteins to be analyzed. The image analysis function in ImageSXM scans the image and selects all of the highlighted proteins within the density slice. Analysis of each protein within the slice is then performed. In addition to height and area information, which permit the calculation of the protein volume, ImageSXM calculates the major and minor axes by fitting the cross-section of each protein to an ellipse. As discussed below, these latter values can be used to remove errors from the data set. The volume for each protein, V_i, is calculated by multiplying the area, A_i, by the corrected average height (total average height, M_i, minus surface height, S)

$$V_i = A_i(M_i - S) \tag{1}$$

Data Filtering. At high surface coverage, the density slice utility cannot distinguish between dimers and two proteins that are close to one another. The image analysis software highlights these closely positioned proteins as a single protein. Visual inspection of these proteins clearly shows that they are two single proteins on the surface. Because ImageSXM fits the protein shape to an ellipse and provides the major and minor axes, the data can be corrected easily by limiting the value of the ratio of the major axis to minor axis. For example, a major:minor axis ratio >2 indicates that the proteins are not associated. A histogram of the corrected volume data is generated for each image.

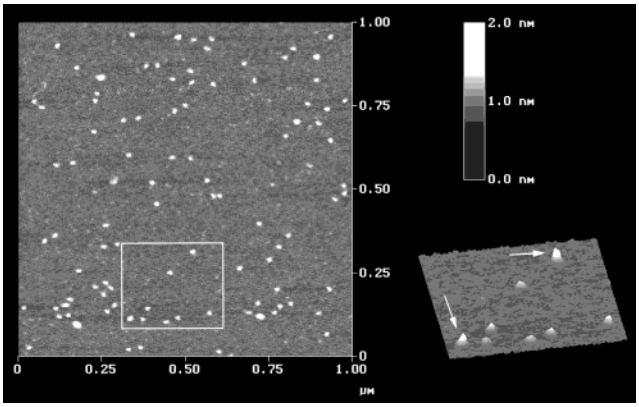

Figure 1. AFM Image of UvrD at 50 nM. The image shows proper surface coverage for volume analysis. A deposition time of 30 s was used for this concentration of UvrD. The surface plot (inset) represents the rectangle area within the image. Arrows in the inset point to dimers; the other proteins within the inset are monomers.

To assess the effect of filtering, histograms are generated using several different "cutoff" values for the minor:major axis ratio. Generally, the number of bins used in each histogram is approximately the square root of the protein population.

Determination of Association Constants. To determine the dimerization constant for UvrD, the protein was deposited at different concentrations, a set of images was collected at each concentration, and each set was subjected to volume analysis. A histogram plot of the filtered protein volumes was generated for each concentration. From the volume histogram, the fraction of dimers is determined by counting the number of proteins that have volumes consistent with the molecular weights of the monomer and dimer species (see Results and Discussion). The dimerization constant is then calculated by analyzing data from depositions at several different protein concentrations and using eq 2,

$$(f/2(1-f)^2) = K_a c \tag{2}$$

where K_a is the association constant ($K_a = [\mathrm{UvrD_2}]/[\mathrm{UvrD}]^2$), c is the concentration of protein monomers, and f is the fraction of monomers as dimers. Equation 2 applies to protein homodimers and is analogous to the method used to determine association constants for nucleic acid duplexes from melting curves.[28] An analogous equation exist for protein heterodimers.[28] Because the dissociation constant, K_d, is equal to the inverse of the K_a, a plot of $f/2(1-f)^2$ versus c (Figure 4) yields the K_d as the inverse slope of the line.

Results and Discussion

AFM Volume Depends Linearly on Molecular Weight. A series of protein molecular weight standards was imaged and the protein volumes determined. Figure 1 shows an ideal surface coverage for volume analysis. Surface coverage should be low enough so that the protein does not coat the entire surface but high enough so that volume analysis may be performed quickly. The measured AFM volumes, calculated using eq 1, of each of the proteins were distributed in a Gaussian fashion. The volume and its uncertainty for a given protein were taken to be the average and standard deviation of the distribution, respectively. Figure 2 shows a plot of the volume of the protein standards versus their molecular weights. The data are fit well by a straight line, indicating that the AFM volume of a protein depends linearly on its molecular weight, as has been observed previously.[23,24] This linear relationship is not highly dependent on the particular tip employed, because these images were collected with several different cantilevers; however, all of the cantilevers

(25) Modrich, P. *Science* **1994**, *266*, 1959−1960.
(26) Sancar, A. *Science* **1994**, *266*, 1954−1956.
(27) Barrett, S. ImageSXM software, Version 1.62, 1999.
(28) Cantor, C. R.; Schimmel, P. R.In *Biophysical Chemistry*; Bartlett, A. C., Ed.; W. H. Freeman and Company: San Francisco, 1980; pp 1183−1264.

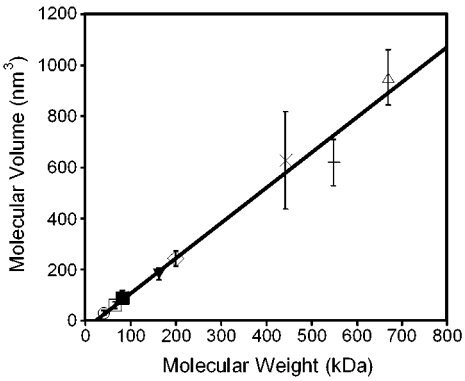

Figure 2. Plot of calculated protein volume versus molecular weight: Alcohol dehydrogenase (○, 83), Bovine serum albumin (□, 97), UvrD monomer (■, 1678), UvrD dimer (▼, 178), β-amylose (◇, 73), RNA polymerase (×, 44), Apoferritin (−, 344), thyroglobulin (△, 37). The number within the parentheses represents the number of proteins analyzed. The line represents the weighted least-squares fit of the data, which is described by the following equation: $V = 1.31(\mathrm{MW}) - 25.0$, where V is volume and MW is molecular weight. ($R^2 = 0.990$.) The error bars represent the standard deviation of the distribution.

were from the same supplier, and therefore the tip geometries probably do not vary significantly. This linear dependence is not necessarily expected because the AFM image is a convolution of the tip geometry and the surface topography as well as surface−tip interactions. It may result from fortuitous compensating effects of imaging and image processing. Specifically, the protein height measured by AFM is generally less than the actual height due, in part, to the flattening procedure, while the protein width is overestimated due to the finite tip size.[3] Irrespective of the underlying causes, the linear dependence of AFM volume on protein molecular weight is robust, especially for proteins with a molecular weight less than 200 000 (Figure 2). This dependence indicates that AFM volumes can be used to reliably estimate the molecular weights and determine association states of proteins.

Previous studies have shown a linear relationship between the measured volume and the molecular weight of proteins.[23,24] Schneider et al. showed this linear relationship using eight proteins with molecular weights ranging from 38 000 to 900 000.[23] The method employed in their study was time-consuming, so only a few molecules for each protein were measured. The number of molecules measured for each molecular weight standard ranged from 6 to 22 proteins, making these data sets to small to do reliable statistical analyses. In addition, they modeled the proteins as a segment of a sphere on a surface to determine the volume, making their technique inapplicable to proteins that are not globular. In our analysis method, height and areas are measured directly and no models are assumed, making the volume measurements relatively insensitive to the geometry of the proteins. In a second study, Bustamante and co-workers used a volume analysis method similar to the one employed here to distinguish dimers and tetramers of NtrC.[24] In their study (unlike Schneider's), they measured a large population of molecules for each protein, but only analyzed three proteins to generate a standard curve. Our method for volume analysis allows for a large population of proteins to be analyzed quickly, and a linear dependence on protein volume and molecular weight was found for many different proteins ranging from 41 000 to 670 000. In addition to this study, four additional proteins and protein−protein complexes of known molecular weight have been investigated by several different individuals, and their volumes were correctly (within 10% of the known value) predicted from the standard curve in Figure 2. (unpublished data) Volume analysis of one of these proteins showed two distributions: one consistent with the molecular weight of the trimer and the other consistent with the molecular weight of the hexamer. The trimer−hexamer equilibrium indicated by the volume analysis data is collaborated by gel filtration and AUC studies.[29] These results further demonstrate that the linear fit is quantitative and that it is not very sensitive to the user, tip employed, or the particular protein being analyzed. It is possible, however, that the slope and intercept of the line may depend on the brand of tips or AFM instrument employed. Consequently, a standard curve using at least a few proteins should be generated prior to using this technique for determining the oligomerization states of proteins. For consistent volume analyses, imaging parameters must remain constant. Unlike optical microscopy techniques, the AFM obtains an image through direct physical contact with the sample; a small tip attached to a cantilever traces the topography of the sample. Because AFM images result from physical interactions, forces between the tip and sample must be kept at a minimum and consistent. For the oscillating mode in air, the drive and clipping amplitudes should be set as low as possible to minimize normal forces exerted on the sample. This procedure was especially important for imaging the proteins used in the standard curve in Figure 2. High forces exerted by the tip distort and compress soft samples such as proteins, resulting in a marked decrease in the protein's volume (data not shown). Large proteins generally are more sensitive to imaging parameters. The reason for this sensitivity probably results from both a greater interaction between the tip and protein and the greater conformational flexibility of larger proteins. This dependence between the measured volume and imaging parameters may contribute to the greater uncertainty in the volume of the three largest proteins in Figure 2.

Dissociation Constant for UvrD. To determine the dimerization constant for UvrD, images of UvrD deposited at different concentrations were collected and volume analysis was performed for each set of images. Prior to analysis, the deposition time that produced optimal surface coverage was determined for each protein concentration. The data were filtered to remove false positives of dimers using major:minor axis ratio cutoff values ranging form 1.5 to 2.0. Histogram plots of the volume of the proteins were generated for each cutoff value. A typical histogram of the molecular volumes for UvrD using a cutoff value of 1.5 is shown in Figure 3. For determining the oligomerization state of a protein, the two distributions should not overlap significantly with each other. The histogram clearly shows two separate distributions: one consistent with the molecular weight of the monomer and the other consistent with the molecular weight of the dimer (Figure 2). Both distributions have a Gaussian shape, allowing the fraction of dimers to be determined by counting the number of proteins under each curve. Generally, a protein population of 300 to 1000 is sufficient to produce a reliable measure of the K_d. The K_d is then calculated using eq 2. Using these data and eq 2, the dimerization constant of UvrD was determined from the graph in Figure 4. The K_d ranges from 1.4 μM at a major:minor axis ratio cutoff value of 1.5 to 0.7 μM at a major:minor axis ratio cutoff value of 2.0. These results are in good agreement with a K_d value of 3.4 μM determined under the same solution conditions using AUC,[30] indicating that the AFM volume

(29) Xue, Y.; Erie, D. A.; Redinbo, M. R. Manuscript in preparation.
(30) Mechanic, L. E.; Hall, M. C.; Matson, S. W. *J. Biol. Chem.* **1999**, *274*, 12488−12498.

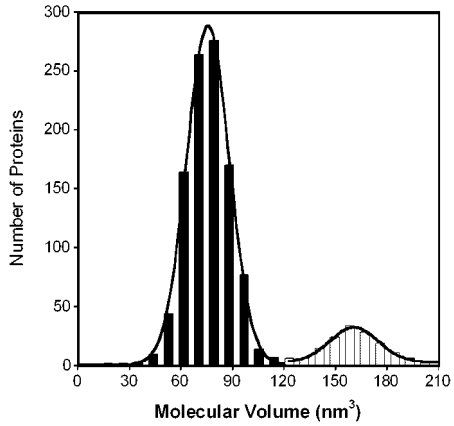

Figure 3. Gaussian fit of the volume histogram for UvrD (250 nM). The solid lines are Gaussian fits of the volume data for monomers (solid bars) and dimers (hatched bars). Each species was fit independently. The number of proteins under each curve represents that species' population.

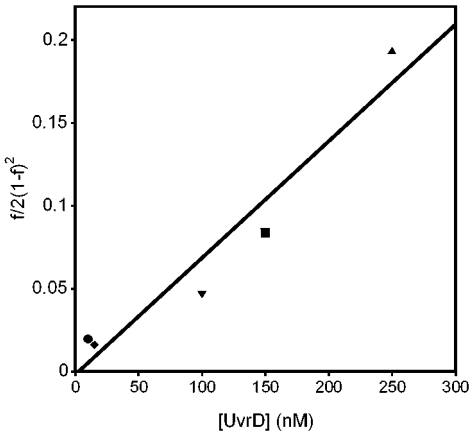

Figure 4. Representative plot used for K_d determination. The data set represents UvrD data using a major:minor axis ratio cutoff value of 1.5. K_d is calculated to be 1.4 μM. [UvrD]: 10 (●), 15 (◆), 100 (▼), 150 (■), and 250 nM (▲). The K_d is the inverse slope of the line and is 1.4 μM for the data shown and decreases to 0.7 μM if the major:minor axis ratio cutoff value is increased to 2.

analysis is a reliable method for determining protein−protein association constants.

Factors that affect the surface deposition of proteins may influence the measured association constant. Two factors are interaction of the proteins with the surface and diffusional properties of the proteins. First, if protein deposition occurs by an equilibrium process, surface interactions with these proteins could perturb the protein−protein association equilibria. This effect can be tested by using surfaces with different properties, such as silicon or graphite. In general, this factor is not a problem in our studies because on mica, most proteins bind irreversibly over the time scale of the deposition. Because the binding is irreversible, the association equilibrium should not be perturbed and, therefore, the population of molecules on the surface should accurately represent the solution population. Second, if the oligomerized protein diffuses significantly more slowly than the monomer, its population could be underrepresented in an AFM image if very short deposition times are used. This effect would decrease the apparent association constant. Fortunately, diffusional factors should only become important for higher order oligomers, because the diffusion coefficients of globular proteins depend inversely on the cube root of their molecular weights.[31] The agreement between our results and those from AUC[30] indicates that these factors do not play a significant role in our experiments.[6]

To our knowledge, this is the first time a single-molecule technique has been used to determine protein−protein association constants. Earlier studies used volume analyses to determine the oligomerization state of proteins[24] but no studies have used AFM to determine an association constant. This study expands the utility of the AFM as a quantitative tool for studying biological systems. This technique can also be used to study the assembly state of proteins that cannot be studied using other techniques. Often it is useful to measure protein−protein association constants under a variety of solution conditions to understand how ionic strength and the addition of ligands, such as ATP, alter a protein's oligomerization state. Many proteins, however, do not remain soluble in these solutions at the high concentrations often required for K_d determination. By using the AFM, much lower protein concentrations may be used, thus eliminating the solubility problems associated with other reliable techniques. In addition, ligands such as ATP that have large extinction coefficients in the UV can interfere with techniques that employ UV detection or excitation. Finally, because AFM visualizes all species present, it is possible to detect minor species. For example, if less than 10% of the sample is in the higher association state, it will be evident in a single molecule analysis but not in bulk solution techniques.

Conclusion

Our method of volume determination allows for rapid analysis of a large population of proteins. Previous studies have shown a linear relationship between protein volume and molecular weight; however, these studies either dealt with a small population of molecules of each protein or analyzed only a few proteins.[20−24] We observe a linear relationship between protein volume and molecular weight with both many different proteins and large protein populations (Figure 2). Furthermore, we found the dependence of AFM volume on molecular weight to be insensitive to the particular protein being imaged, the user, or the tip being employed. Because large protein populations can be analyzed using our method, the statistical analysis is more reliable. Other methods for volume determination prevent the analysis of large data sets due to the time-consuming methods used to calculate the protein's volume.[20,23] In addition, because ImageSXM provides elliptical parameters, it is possible to quickly distinguish dimers from proteins that have coincidentally landed next to one another. Using this method we have successfully determined the K_d for the helicase protein *E. coli* UvrD. This method adds to the arsenal of techniques available to measure binding constants. In closing, AFM provides a powerful tool for determining the oligomerization states of proteins, measuring protein−protein association constants, and detecting minor species, which could be important functionally.

Acknowledgment. We thank Professor Steve Matson of the University of North Carolina at Chapel Hill for supplying UvrD helicase. This work was supported by NIH grants GM54136 and ES09895.

Supporting Information Available: Detailed method of data analysis using ImageSXM (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA005750N

(31) Tinoco, I.; Sauer, K.; Wang, J. C. *Physical Chemistry*; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1995; pp 255−322.